Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume 4 Nomor 2 Juli 2024
e-ISSN: 2827-7953, p-ISSN: 2827-8135, Hal 37-44
DOI: https://doi.org/10.55606/juisik.v4i1.760

Sistem Rekomendasi Film Menggunakan Metode K-NN

Mohammad Amir Fanani
Universitas Pembangunan Nasional "Veteran" Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
Korespondensi penulis: amirfanani98@gmail.com

Abstract. *Recommendation systems are at the heart of cutting through the noise of online data, helping users find content that matches their preferences. In this regard, the K-Nearest Neighbors (KNN) method stands out as a promising approach. KNN, as a similarity-based algorithm, utilizes information from nearest neighbors to make predictions or recommendations. This research explores the implementation of KNN in developing a film recommendation system with a focus on increasing the accuracy and relevance of recommendations. Related references, such as the evaluation of recommendation systems by Herlocker et al. (2004) and the concept of collaborative filtering by Resnick and Varian (1997), are the basis for understanding and improving the potential of KNN. By detailing the exploration of this concept, it is hoped that this research will provide a comprehensive and holistic view in the development of a film recommendation system. From testing using the performance test of the KNN method, namely accuracy, recall, and precision with the best value of 45.4% for accuracy, 45.4% for recall, and 100% precision so that the algorithm can be applied in the film recommendation system.*

Keywords: *Accuracy, K-Nearest Neighbors, Precision, Recal, Recommendation systems.*

Abstrak. Sistem rekomendasi menjadi inti dalam mengatasi hiruk-pikuk data daring, membantu pengguna menemukan konten yang sesuai dengan preferensi mereka. Dalam hal ini, metode K-Nearest Neighbors (KNN) menonjol sebagai pendekatan yang menjanjikan. KNN, sebagai algoritma berbasis kesamaan, memanfaatkan informasi dari tetangga terdekat untuk membuat prediksi atau rekomendasi. Penelitian ini mengeksplorasi implementasi KNN dalam pengembangan sistem rekomendasi film dengan fokus pada peningkatan akurasi dan relevansi rekomendasi. Referensi terkait, seperti evaluasi sistem rekomendasi oleh Herlocker et al. (2004) dan konsep collaborative filtering oleh Resnick dan Varian (1997), menjadi dasar untuk memahami dan meningkatkan potensi KNN. Dengan merinci eksplorasi konsep ini, diharapkan penelitian ini memberikan pandangan menyeluruh dan holistik dalam pengembangan sistem rekomendasi film. Dari pengujian menggunakan uji performa dari metode KNN yaitu accuracy, recall, dan precision dengan nilai terbaik sebesar 45,4% untuk accuracy, 45,4% untuk recall, dan 100% precision sehingga algoritma tersebut dapat diaplikasikan dalam sistem rekomendasi film tersebut.

Kata kunci: Accuracy, K-Nearest Neighbors, Precision, Recall, Sistem Rekomendasi

LATAR BELAKANG

Dalam era digital yang semakin berkembang, pertumbuhan eksponensial data yang tersedia di berbagai platform online menantang pengguna untuk menemukan konten yang relevan dan bermakna. Sistem rekomendasi menjadi solusi utama untuk mengatasi masalah ini, memfasilitasi proses penemuan konten dengan memberikan rekomendasi yang sesuai dengan preferensi pengguna. Salah satu metode yang telah memperoleh perhatian substansial dalam pengembangan sistem rekomendasi adalah metode K-Nearest Neighbors (KNN).

Metode K-Nearest Neighbors (KNN) dikenal sebagai pendekatan berbasis kesamaan yang memanfaatkan informasi dari tetangga terdekat untuk membuat prediksi. Dalam konteks sistem rekomendasi film, KNN dapat diimplementasikan untuk memprediksi preferensi

Received: Desember 09, 2023; Accepted: Januari 09, 2024; Published: Maret 31,2024
** Mohammad Amir Fanani amirfanani98@gmail.com*

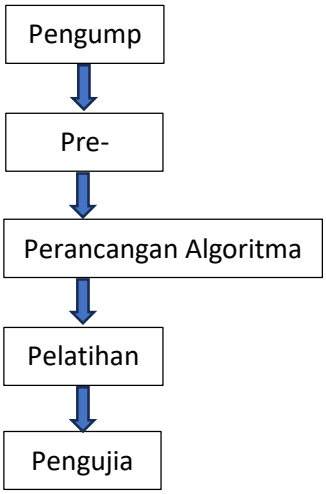
pengguna dengan membandingkan kesamaan profil pengguna dengan pengguna lain dalam ruang atribut film. Analisis historis terkait penilaian film oleh pengguna dapat menjadi dasar untuk mengidentifikasi pola dan merekomendasikan film yang kemungkinan besar akan disukai.

Meskipun KNN telah berhasil digunakan dalam berbagai bidang, penggunaannya dalam konteks sistem rekomendasi masih memerlukan eksplorasi lebih lanjut untuk memahami potensi dan keterbatasannya. Penelitian ini bertujuan untuk menyelidiki implementasi K-Nearest Neighbors dalam pengembangan sistem rekomendasi film, fokus pada peningkatan akurasi dan relevansi rekomendasi.

Penelitian ini merujuk pada konsep-konsep penting dalam evaluasi sistem rekomendasi, perkembangan metode KNN, dan penelitian terdahulu yang relevan dalam pengembangan sistem rekomendasi. Dengan menggali lebih dalam implementasi KNN, diharapkan penelitian ini dapat memberikan pandangan yang lebih holistik dan menyeluruh dalam pengembangan sistem rekomendasi film.

METODE PENELITIAN

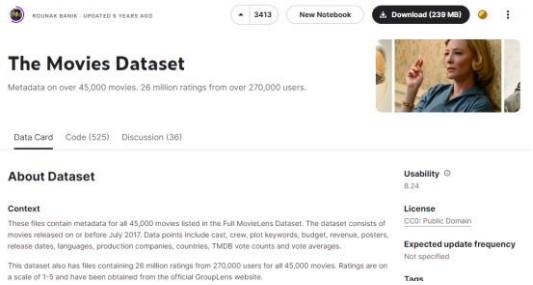
Penelitian ini dimulai dengan melakukan pengumpulan data. Lalu sebelum diproses data tersebut dilakukan beberapa pre-processing data yang berguna agar penelitian nanti minim terjadinya galat. Setelah dilakukan beberapa proses sebelumnya, data sudah dapat digunakan dengan menggunakan Library Surprise yang di dalamnya terdapat Algoritma K-Nearest Neighbors untuk melakukan perhitungan akurasi dan nilai akurasi. Metode penelitian kali ini adalah: (1) Pengumpulan data; (2) Pre-processing data; (3) Perancangan Algoritma; (4) Pelatihan model; (5) Pengujian model. Tahapan-tahapan tersebut disusun dalam gambar 1 di bawah ini.



Gambar. 1 Tahapan Penelitian

1) Pengumpulan Data

Tahapan penelitian yang pertama dilakukan yaitu mengumpulkan data dari The Movies Dataset. Data tersebut diperoleh dari website kaggle.com. Data yang digunakan sebesar 100.004 baris data yang berupa Reviews dari Movies yang pernah ditonton User. Berikut adalah gambar dari sumber data peneliti pada gambar 2.



Gambar. 2 Sumber data

2) Pre-processing Data

Tahap selanjutnya sebelum data tersebut dilakukan Pre-Processing Data di mana data tersebut dicari dan diperbaiki lagi sebelum benar-benar diproses oleh Algoritma. Data yang dicari berupa data yang kosong, data yang belum sesuai maupun data yang sudah rusak sehingga apabila tetap dilakukan proses Algoritma akan terjadi galat yang bisa menghambat bahkan tidak memunculkan data yang ingin diperoleh dari metode ini. Cara ini disebut dengan Cleaning data yang berarti membersihkan data yang masih mentah sebelum dipakai. Prosesnya bisa dengan menghapus atau mengganti data yang kurang sesuai. Setelah data dibersihkan, ditentukan proses klasifikasi data dengan menentukan nilai K pada Algoritma yang akan dipakai. Peneliti menentukan nilai K yaitu sebesar k=5, k=8, dan k=10 pada penelitian ini. Hasil yang didapatkan akan diuji performanya dengan menggunakan pengukuran nilai akurasi, nilai presisi dan nilai recall.

3) Perancangan Algoritma

Algoritma yang akan digunakan Peneliti adalah K-Nearest Neighbor. Algoritma ini digunakan peneliti karena dapat menentukan nilai rekomendasi berdasarkan jarak terdekat dari data yang ingin dicari. Penentuan jarak dari metode K-Nearest Neighbor dihitung menggunakan beberapa cara. Pelatihan dan pengujian dilakukan agar peneliti mendapatkan nilai terbaik dari setiap perhitungan untuk Algoritma K-Nearest Neighbor. Maka pada perancangan algoritma ini disusunlah suatu tahapan yaitu:

- Menentukan parameter k (jumlah tetangga terdekat).
- Menentukan persentase data yang akan digunakan pelatihan dan pengujian data.
- Melakukan perhitungan menggunakan beberapa perhitungan yang tersedia.
- Sortir data dari nilai k data terkecil.

4) Pelatihan

Pada penelitian kali ini, peneliti melakukan pelatihan pada dataset untuk melatih Algoritma K-Nearest Neighbor. Pelatihan tersebut dilakukan pada The Movies Dataset yang sudah melalui tahap pre-processing dan sudah dilakukan perhitungan jarak K-Nearest Neighbor. Data yang akan diproses akan dibagi menjadi 2 yaitu data training atau pelatihan sebanyak 70% dan 80%, serta data testing atau data tes sebanyak 30% dan 20%. Data pelatihan yang akan dipakai pada persentase 70% adalah sebanyak 70.002 data dan pada 80% adalah sebanyak 80.003 data dari dataset.

5) Pengujian

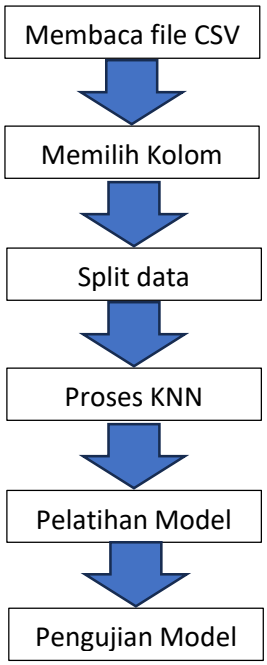
Pengujian dilakukan menggunakan Algoritma K-Nearest Neighbor. Pengujian ini bertujuan untuk mencari nilai performa dari Algoritma K-Nearest Neighbor dengan menggunakan beberapa perhitungan uji coba yaitu accuracy, precision, dan recall. Dalam pengujian kali ini membuktikan nilai akurasi pada Algoritma K-Nearest Neighbor. Pengujian kali ini menggunakan data sebanyak 30% atau 30.002 dan 20% atau 20.001 data pada Algoritma K-Nearest Neighbor.

HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data sebanyak 100.004 data untuk diproses klasifikasi dalam sistem rekomendasi. Library yang digunakan untuk memproses data tersebut adalah surprise library yang berisi beberapa macam metode perhitungan dalam K-Nearest Neighbor.

1) Model KNN

Sistem rekomendasi yang digunakan dalam memproses The Movies Dataset adalah K-Nearest Neighbor dengan menggunakan library surprise dengan dilakukan beberapa perhitungan meliputi cosine, mean squared difference, pearson, dan pearson baseline. Proses untuk membentuk model KNN dilalui beberapa proses yang akan dijelaskan pada gambar 3.



Gambar. 3 Alur Model KNN

Pada gambar 3, beberapa tahapan yang dilakukan untuk klasifikasi rating film dengan Algoritma K-Nearest Neighbor sebagai berikut:

a. Membaca file CSV yang mana file tersebut mengandung data yang akan digunakan peneliti di antaranya rating dan UserId.
b. Memilih kolom dari file CSV sebelumnya dengan memilih dua di antara 4 kolom yang akan digunakan.
c. Membagi isi data yang akan digunakan dengan persentase 70% untuk data pelatihan dan 30% untuk data uji. Serta model lain yaitu dengan persentase 80% data pelatihan dan 20% untuk data uji.
d. Dilakukan pelatihan pada data yang sudah dibagi dengan metode Cosine, MSD, Pearson dan Pearson Baseline.
e. Dilakukan Pengujian menggunakan metode Accuracy, Recall, Precision

2) Performa KNN

Dari hasil klasifikasi rating film dengan metode K-Nearest Neighbor dengan data yang berjumlah 100.004 data. Data-data tersebut akan dibagi menjadi 2 bagian untuk pelatihan dan pengujian sebanyak 70% untuk pelatihan dan 30% untuk pengujian, serta 80% untuk pelatihan dan 20% untuk pengujian. Pengujian klasifikasi tersebut akan dibagi menjadi beberapa parameter dari banyaknya K-Nearest Neighbor yaitu dengan nilai k=5, k=8, dan k=10. Berikut hasil dari pengujian pada parameter dari k=5 dan data tes 30%

Tabel 1

Hasil Akurasi K=5 Data tes 30%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	41,93%	41,93%	100%
2	MSD	42,53%	42,53%	100%
3	Pearson	42,81%	42,81%	100%
4	Pearson Baseline	43,87%	43,87%	100%

Berikutnya adalah parameter k=5 dan data tes 20%.

Tabel 2

Hasil Akurasi K=5 Data tes 20%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	42,63%	42,63%	100%
2	MSD	43,47%	43,47%	100%
3	Pearson	42,93%	42,93%	100%
4	Pearson Baseline	44,4%	44,4%	100%

Selanjutnya dari parameter k=8 dan data tes 30%

Tabel 3

Hasil Akurasi K=8 Data tes 30%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	43,12%	43,12%	100%
2	MSD	43,64%	43,64%	100%
3	Pearson	43,86%	43,86%	100%
4	Pearson Baseline	44,65%	44,65%	100%

Dan dari parameter k=8 dengan data tes 20%

Tabel 4

Hasil Akurasi K=8 Data tes 20%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	43,47%	43,47%	100%
2	MSD	44,33%	44,33%	100%
3	Pearson	44,56%	44,56%	100%
4	Pearson Baseline	45,22%	45,22%	100%

Lalu parameter k=10 dengan data tes 30%

Tabel 5

Hasil Akurasi K=10 Data tes 30%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	43,49%	43,49%	100%
2	MSD	43,97%	43,97%	100%
3	Pearson	44,24%	44,24%	100%
4	Pearson Baseline	44,83%	44,83%	100%

Dan yang terakhir dari parameter k=10 dengan data tes 20%

Tabel 6

Hasil Akurasi K=10 Data tes 20%

No.	Accuracy Based	Accuracy	Recall	Precision
1	Cosine	43,94%	43,94%	100%
2	MSD	44,69%	44,69%	100%
3	Pearson	44,7%	44,7%	100%
4	Pearson Baseline	45,4%	45,4%	100%

Berdasarkan pengujian yang sudah dilakukan pada masing-masing parameter dari k=5 sampai dengan k=10 dengan data uji 20% dan 30% pada masing-masing parameter ditemukan nilai accuracy tertinggi yaitu pada parameter k=10 dengan data tes 20% menggunakan metode pearson baseline sebesar 45,4%. Kemudian nilai recall tertinggi ditemukan pada parameter k=10 dengan data tes 20% menggunakan metode pearson baseline sebesar 45,4%. Dan nilai precision tertinggi pada semua pengujian yang dilakukan yaitu 100%. Nilai tersebut apabila dibentuk menjadi tabel sebagai berikut

Tabel 7

Hasil Akurasi Terbaik

No.	Jenis Performa	Nilai
1	Accuracy	45,4%
2	Recall	45,4%
3	Precision	100%

Berdasarkan tabel sebelumnya ditemukan nilai tertinggi dari masing-masing pengujian di antaranya nilai accuracy sebesar 45,4%, nilai recall 45,4% dan precision sebesar 100% pada pengujian metode pearson baseline dengan data uji 20%.

KESIMPULAN DAN SARAN

Pada penelitian sistem rekomendasi pada film menggunakan algoritma K-Nearest Neighbors. Hasil pengujian akurasi akurasi performa dari model algoritma K-nearest Neighbors yang dirancang sebelumnya dengan hasil accuracy 45,4%, recall 45,4%, dan precision 100%. Sistem rekomendasi pada film menggunakan metode algoritma K-Nearest Neighbors untuk mengklasifikasikan rating film yang dapat diterapkan dan diimplementasikan pada User untuk merekomendasikan film tertentu kepada User yang lain, sehingga pada penelitian selanjutnya dapat dikembangkan lebih lanjut dengan menggabungkan beberapa metode algoritma untuk sistem rekomendasi film yang lebih baik.

DAFTAR REFERENSI

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1), 5-53.Risdwiyanto, A. & Kurniyati, Y. (2015). Strategi Pemasaran Perguruan Tinggi Swasta di Kabupaten Sleman Yogyakarta Berbasis Rangsangan Pemasaran. *Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship*, *5*(1), 1-23. http://dx.doi.org/10.30588/SOSHUMDIK.v5i1.142

Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295).

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to Usenet news. Communications of the ACM, 40(3), 77-87.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (pp. 43-52).